Cybersecurity for Distributed Science: Fortifying the Front-lines of the Cybersecurity War

Deb Agarwal
DAAgarwal@lbl.gov
Lawrence Berkeley Laboratory

Threats

- Viruses
- Worms
- Malicious software downloads
- Spyware
- Stolen credentials
- Insider Threat
- Denial of service
- Root kits
- Session hijacking
- Agent hijacking
- Man-in-the-middle
- Network spoofing
- Back doors
- Exploitation of buffer overflows and other software flaws
- Phishing
- Audits / Policy / Compliance
- ?????

Example - Credential Theft

- Widespread compromises
  - Over 20++ sites
  - Over 3000+ computers
  - Unknown # of accounts
  - Very similar to unresolved compromises from 2003
- Common Modus Operandi
  - Acquire legitimate username/password via keyboard sniffers and/or trojaned clients and servers
  - Log into system as legitimate user and do reconnaissance
  - Use "off the shelf" rootkits to acquire root
  - Install sniffers and compromise services, modify ssh-keys
  - Leverage data gathered to move to next system
- ***The largest compromises in recent memory (in terms of # hosts and sites)***

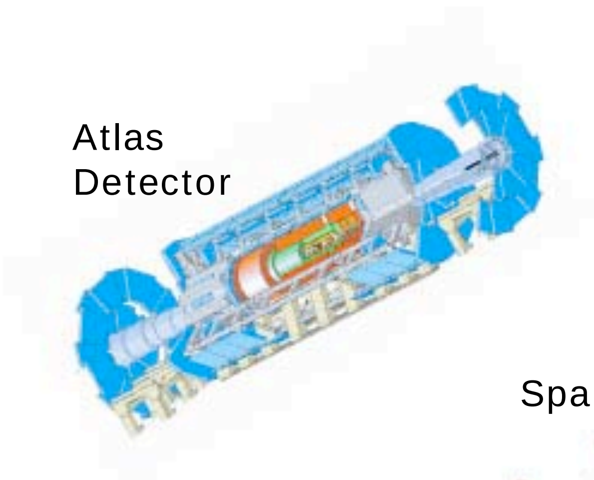
Cybersecurity Trend - Reactive

- Firewall everything – only allow through vetted applications with strong business need
- Users never have administrator privileges
- All software installed by administrators
- ***All systems running automated central configuration management and central protection management***
- ***Background checks for users with physical presence for issuance of HSPD-12 cards (PIV)***
- ***No access from untrusted networks***
- ***Conformance and compliance driven***
- ***It is a war***

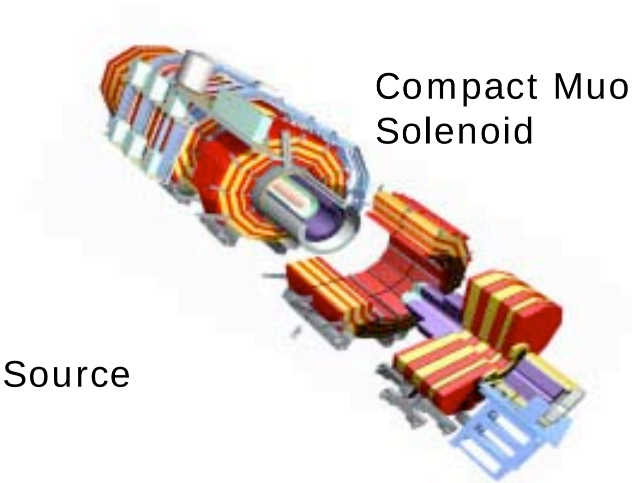
Distributed Science Reality

- Collaborations include as many as 1000's of scientists
- Collaborators located all over the world
- Many users never visit the site
- Virtual organization involved in managing the resources
  - Include multiple sites and countries
  - Distributed data storage
  - Distributed compute resources
  - Shared resources
- Do not control the computers users are accessing resources from
- High performance computing, networking, and data transfers are core capabilities needed
- Authentication, authorization, accounting, monitoring, logging, resource management, etc built into middleware
- ***These new science paradigms rely on robust secure high-performance distributed science infrastructure***

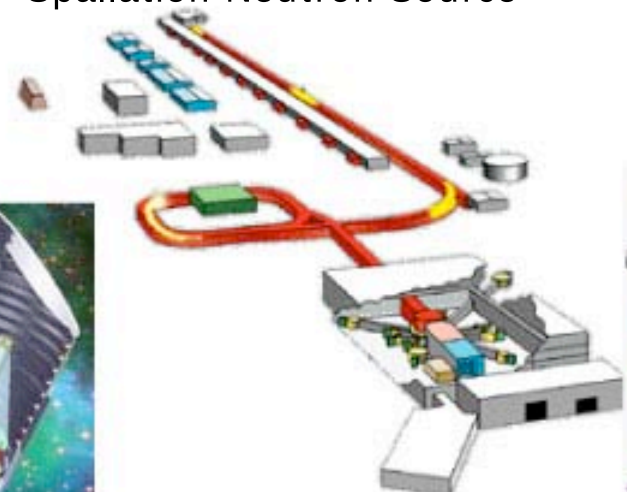
Experiments



Atlas Detector

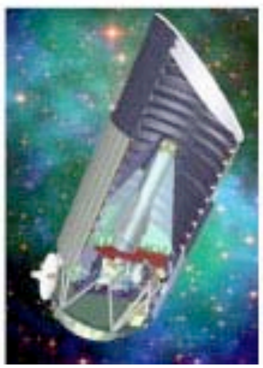


Compact Muon Solenoid

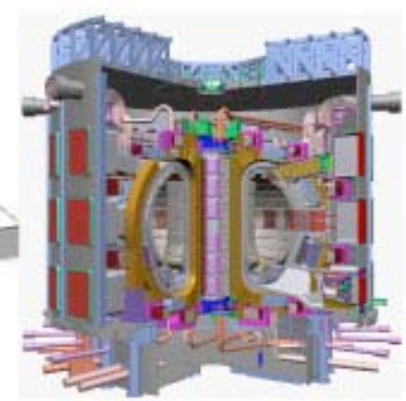


Spallation Neutron Source

Ultrahigh Voltage Electron Microscope

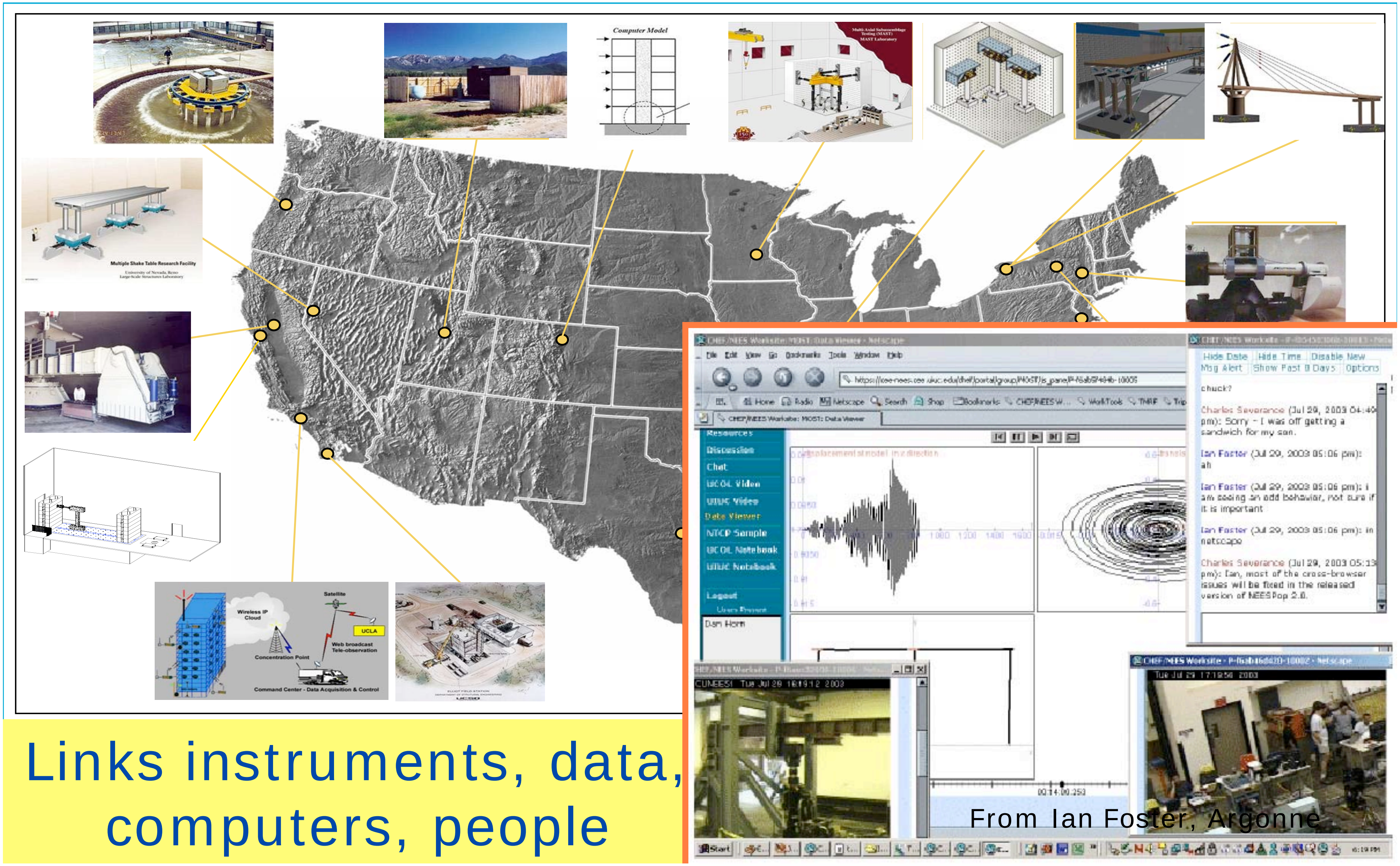


Supernova/ Acceleration Probe



ITER Tokamak

NSF Network for Earthquake Engineering Simulation

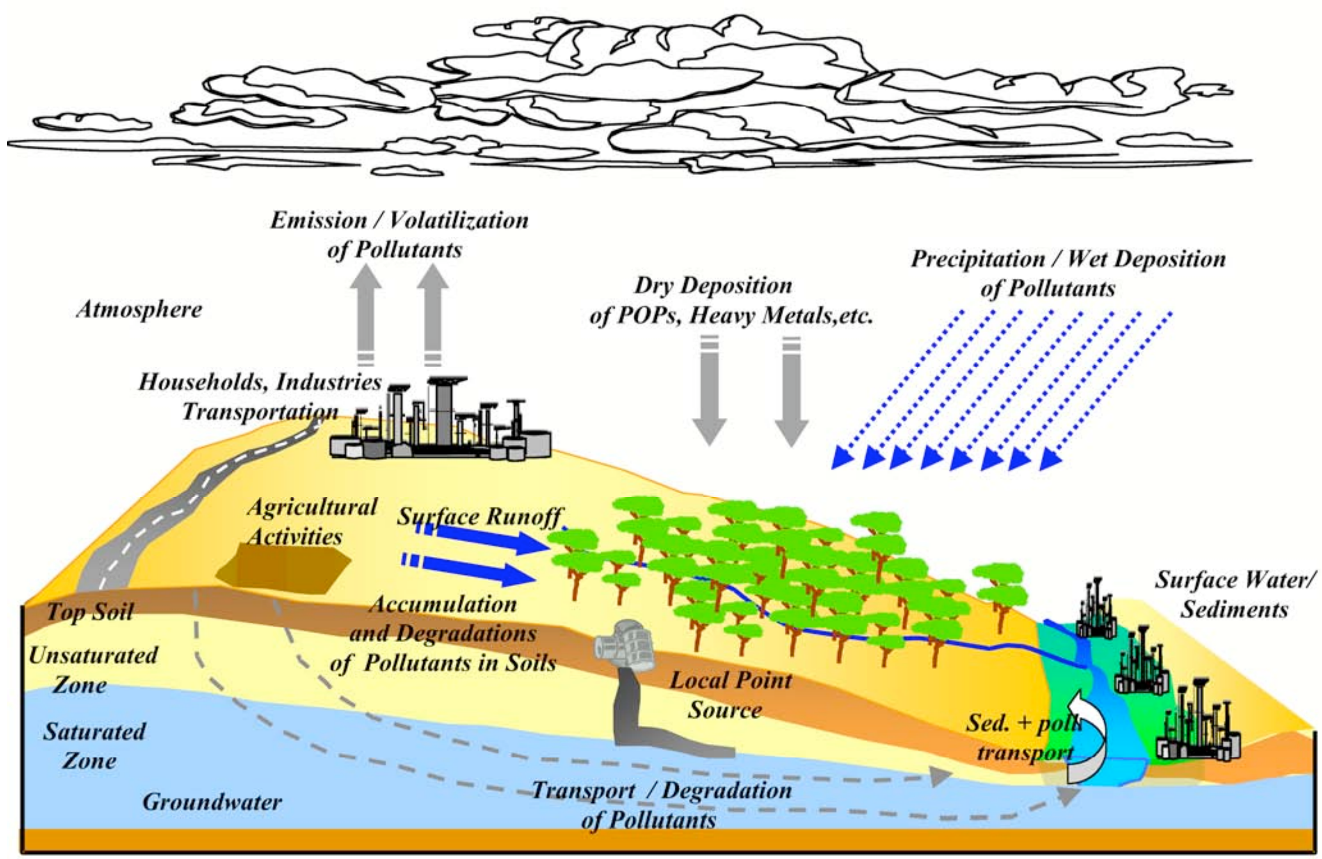


Links instruments, data, computers, people

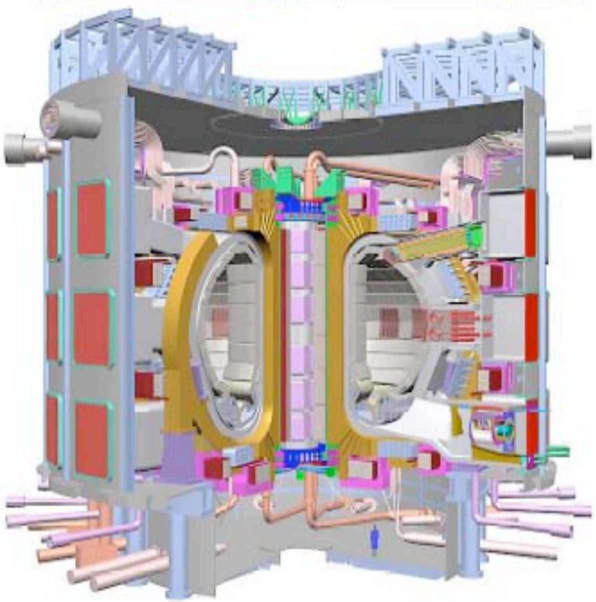
From Ian Foster, Argonne

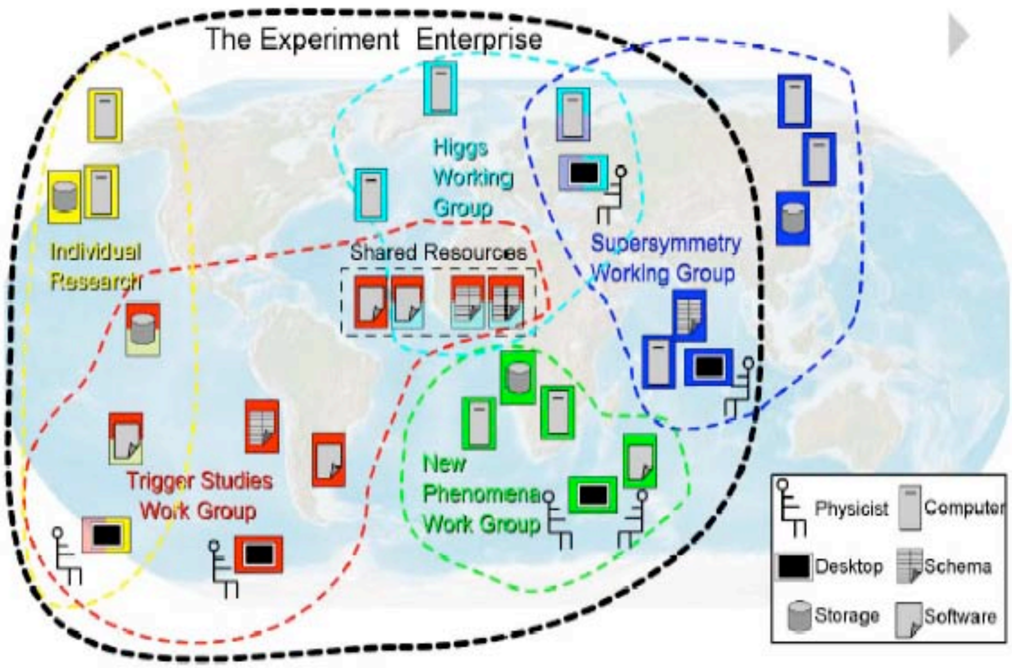
U.S. Department of Energy
Office of Science

Hydrology Synthesis



Science Has Become a Team Sport

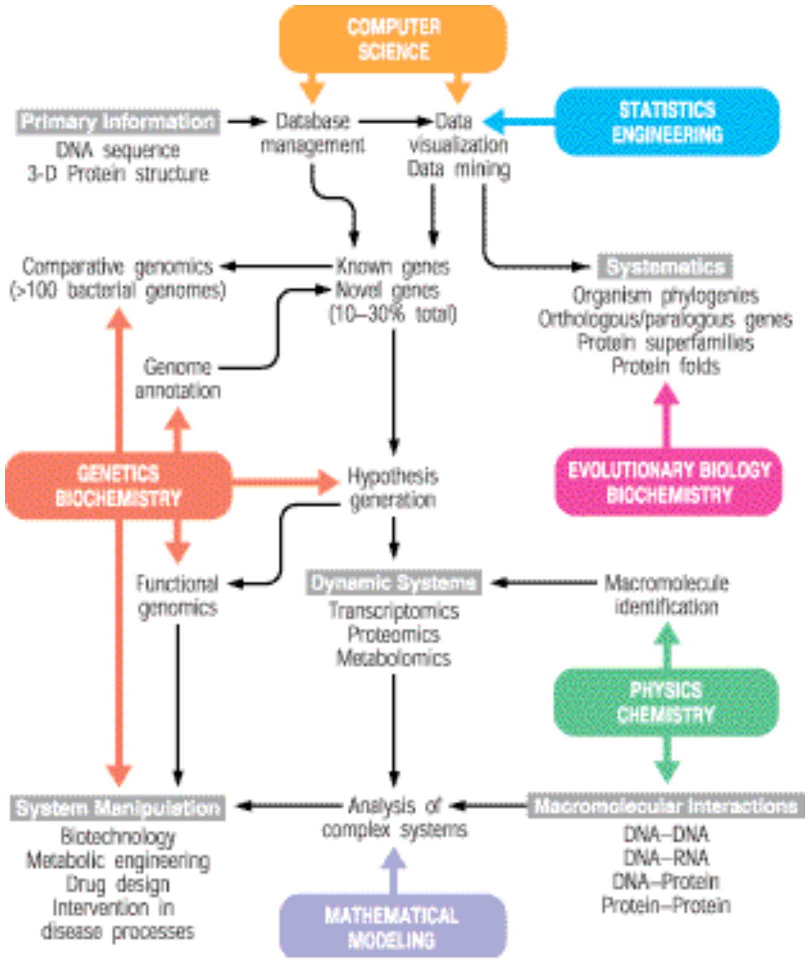




from Dave Schissel, GA

Teams Sharing Data and Expertise

Systems Biology: "studying *biological systems by systematically perturbing them (biologically, genetically or chemically); monitoring the gene, protein, and informational pathway responses; integrating these data; and ultimately formulating mathematical models that describe the structure of the system and its responses to individual perturbations*" *(Ideker et al., 2001 Annu, Rev. Genom. Hum. Genet. 2:343)*



Konopka, 2004 *ASM News* 70:163

from Yuri Gorbi, PNNL

U.S. Department of Energy
Office of Science

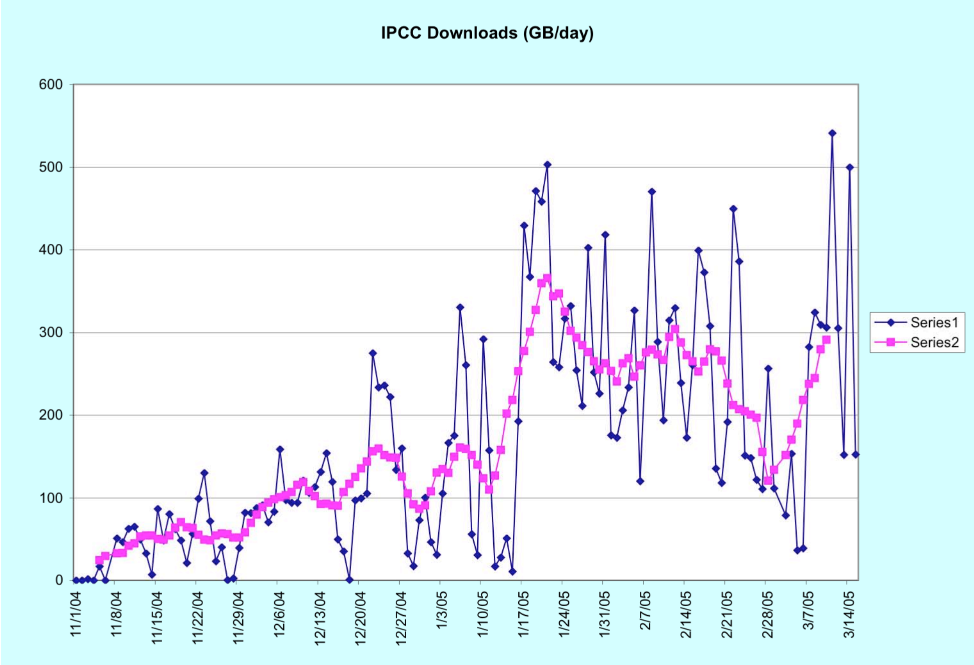
Science Requirements for Networks - 2003

Science Areas	2003 *End2End* Throughput	5 years End2End Throughput	5-10 Years End2End Throughput	Remarks
High Energy Physics	0.5 Gb/s	100 Gb/s	**1000 Gb/s**	high bulk throughput
Climate (Data & Computation)	0.5 Gb/s	160-200 Gb/s	**N x 1000 Gb/s**	high bulk throughput
SNS NanoScience	Not yet started	1 Gb/s	**1000 Gb/s + QoS for control channel**	remote control and time critical throughput
Fusion Energy	0.066 Gb/s (500 MB/s burst)	0.198 Gb/s (500MB/ 20 sec. burst)	**N x 1000 Gb/s**	time critical throughput
Astrophysics	0.013 Gb/s (1 TBy/week)	N*N multicast	**1000 Gb/s**	computational steering and collaborations
Genomics Data & Computation	0.091 Gb/s (1 TBy/day)	100s of users	**1000 Gb/s + QoS for control channel**	high throughput and steering

Delivering Climate Data

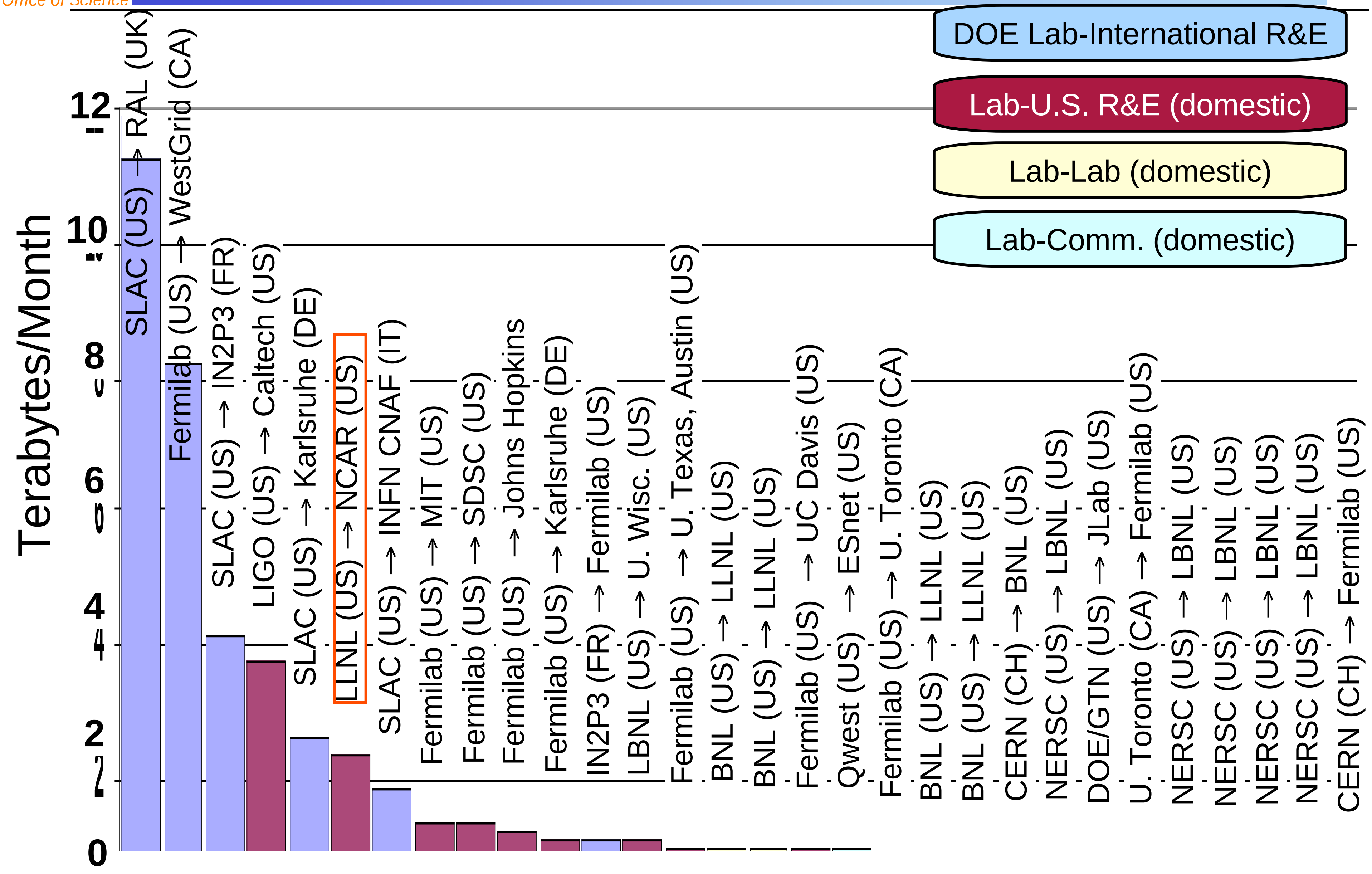
- Earth System Grid (ESG) provides production service (secure portal) to distribute data to the greater climate community.
  - Over 18 terabytes (~40k files) published since December 2004
  - About 300 projects registered to receive data
  - Over 22 terabytes of data downloaded (~125K files) with 300 gigabytes daily.
- Analysis results of IPCC data, distributed via ESG, were presented by 130 scientists at a recent workshop (March 2005).

Enabling Access to Climate Data from the Intergovernmental Panel on Climate Change



from Don Middleton, NCAR

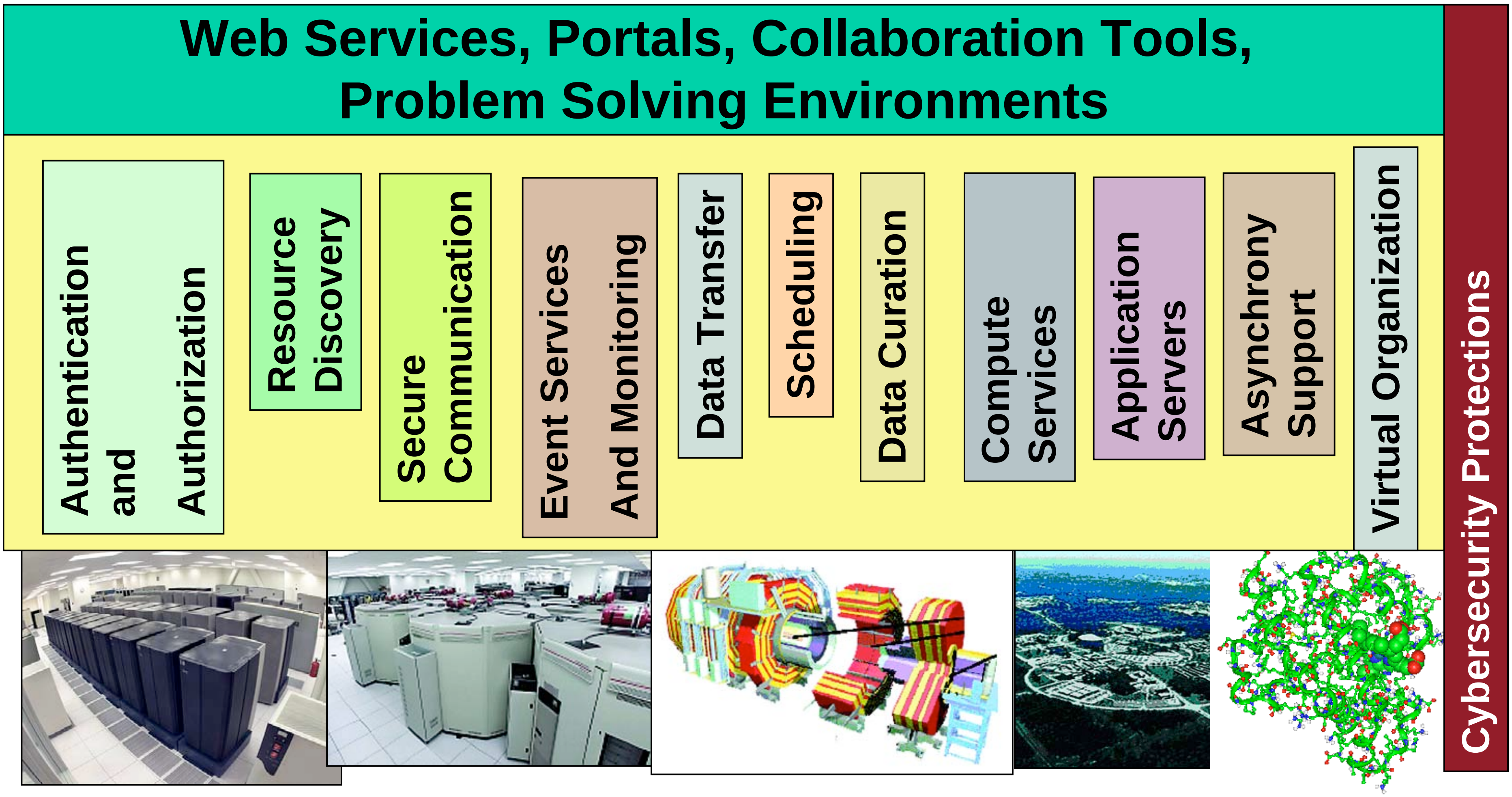
Source and Destination of the Top 30 ESnet Flows, Feb. 2005



Cybersecurity and Infrastructure to Support Distributed Science

- Preserve
  - Access to national user facilities
  - Participation in international collaborations
  - Ability to host scientific databases and repositories
  - Innovation and prototyping capabilities
- Protect
  - High performance computers
  - Experiment systems
  - Desktop and laptop systems
  - Ability to do science
- ***Need to figure out how to preserve and support open science while protecting the resources from cyber incidents***

Robust Science Support Framework



Current Research Middleware Reality wrt Cybersecurity

- Distributed Science Infrastructure is developed independent of operational cybersecurity considerations
  - Implications of site mechanisms
  - Protections from malicious code
  - Vulnerability testing
  - Interoperability with site cybersecurity mechanisms
  - Not commercial software
- Typically there is a long process of debugging prototype deployments
  - Negotiating ports and protocols with each site's cybersecurity group
  - Debugging unexpected behaviors
  - Debugging middleware security mechanisms
  - Identifying causes of performance problems
- ***This is a cross-agency and international issue***

Science is on the Front Lines

- The techniques needed to protect the open science environment today are needed by other environments tomorrow – Past examples
  - Network intrusion detection
  - Insider threat
  - Defense in depth
  - High performance capabilities
- A next set of concerns
  - Reducing credential theft opportunities
  - Detection of insider attacks
  - Communication and coordination between components to recognize and react to attacks in real time
  - Tools which address day zero-1 vulnerabilities
  - Improved analysis techniques – data mining and semantic level searches
  - Prevention and detection of session hi-jacking

Current Operational Reality

- Cybersecurity group
  - Protect border
  - Protect network
  - Some host protections
  - Control access patterns
- System Administrators
  - Protect hosts
  - Authorize users
  - Define access capabilities
- Applications and software
  - Authenticate users
  - Authorize users
  - Open ports/connect to servers/transfer data
- Virtual Organizations
  - Fine-grained authorization
  - Policy enforcement

Protecting High Performance Distributed Science

- Coordination between cybersecurity components
  - Border intrusion detection mechanisms
  - Network intrusion detection mechanisms
  - Host security mechanisms
  - Software authentication and authorization mechanisms
- Authentication mechanisms for users who never physically visit the site
- Analysis of cybersecurity data particularly in high-performance environments
- Efficient forensics information gathering
- Cybersecurity as an integral consideration in building middleware
- Proxy mechanisms
- Continuous data collection and data correlation
- Forensics collection including middleware
- Improved recovery capabilities – it is currently weeks to recover a supercomputer
- Operations, research, and middleware developers teamed
- ***A new operations oriented Cybersecurity R&D effort is needed to help protect open science***

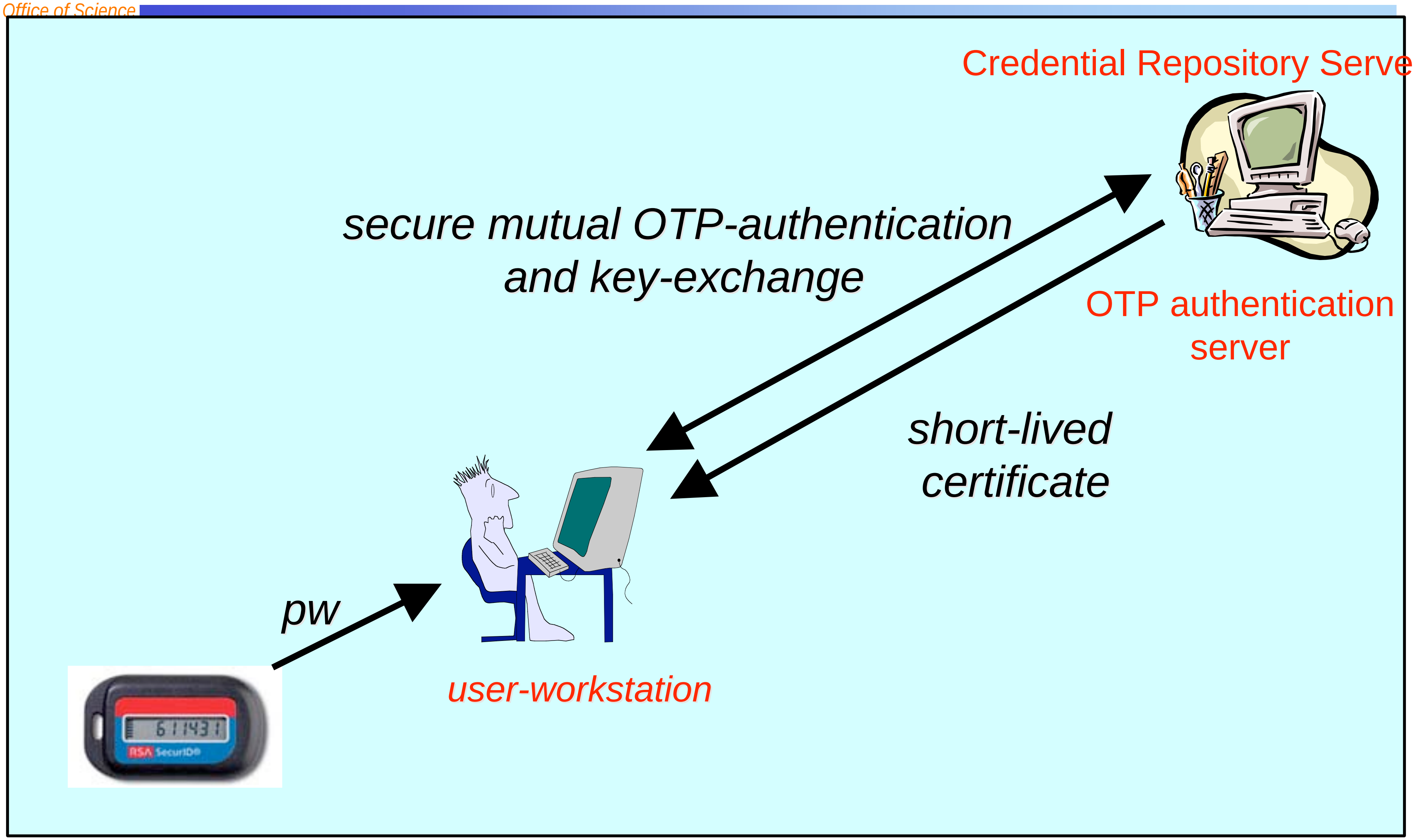
Example Advantages of Research and Operations Working Together

- Bro – network intrusion detection
  - Implemented and deployed through teaming between research and operations
  - Introduced layered approach to high-speed intrusion detection
  - Protocol awareness allowed detection of anomalous behavior at the protocol level
  - Developed policy language and interpreter to describe policy
  - Research platform for investigation of new approaches and events
  - Developments based on experience with real traffic and the operational environment
  - Currently leveraging the Bro communication capabilities to add decryption of encrypted traffic streams

Example2: One-time Password

- Deploying at many sites and facilities to combat credential theft
- Many products out there on the market
- 1-factor, 2-factor, cards, software-based, etc
- Federation an important issue to reduce cost and the number of tokens a user must carry – must be secure to avoid creating cross-site propagation vectors
- Analysis from a cryptographic perspective of the various tools identified important short-comings
- Needs to be integrated with distributed science infrastructure to be fully realized

Using OPKeyX in Grid environments



Conclusions

- Distributed science has become core to the conduct of science
- Robust, **secure**, and supported distributed science infrastructure is needed
- Attackers are getting more malicious and quicker to exploit vulnerabilities
- Need to set the example for protecting distributed infrastructure
- COTS and traditional cybersecurity research are key components of the solution but they do not solve critical aspects of the problem
- ***Need to partner cybersecurity operations, cybersecurity researchers, system administrators, and middleware developers***
- ***Need to rethink cybersecurity for collaborative science***